

An Examination of the Efficacy of the Plagiarism Detection Software Program Turnitin

by

Kevin Walchuk

A thesis submitted in conformity with the requirements
for the Degree of
Master of Education
Graduate Department of Education in the
University of Ontario Institute of Technology

© Copyright by Kevin Walchuk 2016

Abstract

This research project examined the effectiveness of the plagiarism detection program Turnitin in identifying clearly copied, verbatim textual material from peer reviewed and published academic journal articles. It also considered both the effectiveness and appropriateness of tasking plagiarism detection programs, such as Turnitin, with the responsibility of identifying semantic plagiarism, defined for this purpose as the plagiarism of the meaning of text. Results of the exploratory study conducted as part of this research project suggest that while Turnitin is able to match the majority of identical text sampled, it is inconsistent in identifying similar or copied written content in submitted work. Specifically, a gap was found in the program's ability to identify copied textual content in submissions originating from subscription-based journal providers that do not have content agreements with Turnitin. Results also suggest that the original age of submitted publications may affect their likelihood of detection, as older journal articles are less likely to be included, and thus identified, in searchable databases. Furthermore, the study found that coincidental matches, such as the identification of common references, quotations, and commonly used phrases, are regularly produced by Turnitin in its findings of similarities. Finally, the results suggest that Turnitin is able to successfully identify verbatim matching text, but does not effectively assess for semantic plagiarism. The identification, assessment, and interpretation of more advanced plagiarism methods, such as purposed unsourced paraphrasing and the plagiarism of ideas, is outside the purview of its design and function. This research suggests that educators who use or wish to use Turnitin should be made aware of its limitations.

Keywords: plagiarism, plagiarism detection, plagiarism prevention, cheating, Turnitin MyDropBox, SafeAssign, academic honesty, authorship, semantic detection.

Acknowledgements

I would like to extend my sincere thanks to Dr. Jia Li, my supervisor on this research project. Her expertise, dedication, and precision were instrumental in the design and development of this project from start to finish. I would also like to thank Dr. Michael Owen, my co-supervisor, for his insightful comments and perspective, as well as his introduction to enriched literature on the research topic.

Thanks also are extended to Dr. Laura Pinto, Dr. Suzanne de Castell, and Dr. Robin Kay, for their valuable advice and counsel throughout this process. I would also like to thank all of the educators and support staff at the University of Ontario Institute of Technology for their support and assistance throughout my time as a graduate student in the Faculty of Education.

Finally, a special thanks goes to my wife, Amanda Walchuk, and my children, John and Sam, for their incredible patience and support throughout this endeavor.

Contents

Abstract	i
Acknowledgements.....	ii
1 Introduction	1
2 Literature Review	4
2.1 Defining Plagiarism	5
2.2 Forms of Plagiarism	5
2.3 Effectiveness of Plagiarism Detection Software	8
2.4 Plagiarism Detection Software: Pedagogical Application	16
2.5 Summary of the Literature Review.....	18
3 Methods.....	20
3.1 Research Questions	20
3.2 An Exploratory Study Design and Procedure Using Turnitin	20
4 Results	24
4.1 Success in Identifying Textual Similarities and Primary Sources of Copied Written Submission.....	24
4.2 Detecting Plagiarized Ideas and Unsourced Paraphrased Content.....	28
5 Discussion.....	28
5.1 Gaps in Identifying Textual Similarity and Primary Sources	29
6 Conclusion	33
7 Future Research	35
References	38
Appendix A – Results of Exploratory Studies.....	48

1 Introduction

The use of commercial plagiarism detection programs has increased in post-secondary education since their development in the late 1990s (Batane, 2010). This has largely been in response to the perception that plagiarism is a growing problem in post-secondary education and reflects a fear that many students are using the Internet to obtain content, or even complete assignments, and then submitting these works as their own (Dahl, 2007; Evans, 2006; Howard, 2007). In fact, over the past two decades, the education world has been enveloped by what Howard (2007) calls a “sense of impending doom” (p. 3) over the specter of Internet-facilitated plagiarism and the perception that a rise in academic thievery has the potential to undo long-standing traditions in written composition.

For this research, plagiarism will be defined as a process by which someone represents a source’s words or ideas as their own, without crediting the source. This definition will also include semantic plagiarism, defined as the plagiarism of the essential meaning of text, including the intentional plagiarism of ideas. Recently, concern has grown in academic environments over the occurrence of plagiarism, due in part to the increasing availability of source material online and the ease by which unattributed copying of ideas and content can occur (Evans, 2006). The academic community has responded with a variety of strategies aimed at countering this reality or perceived reality. These strategies have included raising awareness and understanding about plagiarism, attempting to clarify what constitutes plagiarism in academic settings, and the introduction of web-based plagiarism detection programs meant to assist students and educators in identifying plagiarism in written work (Dahl, 2007).

Several authors have examined the evolution of web-based plagiarism in the digital age. Kitalong (1998) makes this observation:

The Internet's rich repository of online texts provides an unprecedented opportunity for plagiarism. With a few strokes, anyone with an Internet connection has access to a wealth of easily downloaded material. (p. 255)

Evans (2006) argues that the threat posed to the integrity of academic work by the Internet and its supporting applications is an increasingly serious problem in education:

The ease with which text, numbers and computer codes can be moved between students and institutions has the potential to undermine traditional forms of learning and assessment. (p. 87)

This growing perception of digital media as facilitators of plagiarism has led to an increase in the use of plagiarism detection programs to identify academic dishonesty in post-secondary classrooms (Graham-Matheson & Starr, 2013). For example, iParidigm, the company that developed the plagiarism detection program Turnitin, now claims that more than 15,000 rights-paying member institutions use its services worldwide (Turnitin, 2016). Given their increasing use, the effectiveness of plagiarism detection programs as digital tools aimed at identifying acts of plagiarism in academic settings is worth examining.

Plagiarism detection programs are successful when they are able to identify submitted textual content in searchable databases, including academic and commercial online repositories of written work. Academic research has found inconsistencies in the ability of programs such as Turnitin to identify both verbatim copied textual material and semantic plagiarism, such as the plagiarism of ideas, in submitted academic works originating from a variety of sources, including subscription databases, public web pages, and freely available digital databases such as Google Scholar and Wikipedia (Fiedler & Kaner, 2012; Hill & Page, 2009). Commercial plagiarism detection programs such as Turnitin lack the ability to recognize semantic plagiarism, as the assessment

and interpretation of what constitutes plagiarism and originality in written works falls outside the program's detection abilities. For example, Turnitin is unable to consider, analyze, or reflect on the variety of conventions that contribute to the shaping of written work through the influence of other texts. These include the appropriation of ideas and writing styles, and the shaping of written work through the influence of other source works. This kind of analysis and interpretation is instead left to the program's users to carry out. The following are examples of some of the types of written plagiarism that might require semantic consideration and adjudication outside the purview and function of Turnitin:

- The plagiarism of another author's ideas, concepts, or opinions outside the realm of common knowledge and without attribution.
- Unsourced paraphrasing methods meant to avoid plagiarism detection, including the adding or removing of words and characters; adding deliberate spelling and grammar mistakes; inserting synonyms; the use of alternative syntactic forms for the same expression; patchwriting, which involves writing passages that are not copied exactly but have nevertheless been borrowed from another source and changed (Howard, 1995; Kakkonen & Mozgovoy, 2010).
- Academic work that has been collaboratively created by a group but is represented as if it were a student's own (Park, 2004).
- The intentional plagiarizing of translated texts that are unsupported by the acknowledgement of the original source (Kakkonen & Mozgovoy, 2010).
- The use of textual creations produced by an independent ghostwriter. This could include work that has been created by a single person or a group and represented as if it were a student's own (Park, 2004).

- The presentation/representation of someone else's work in a different medium without attribution (Kakkonen & Mozgovoy, 2010).
- The reproduction of the same or almost identical work for more than one purpose, such as submitting the same assignment for multiple courses (Park, 2004).
- The deliberate and inaccurate use of references, including the use of made-up references or the linking of references to incorrect or non-existent sources (Kakkonen & Mozgovoy, 2010).

Plagiarism detection program providers such as Turnitin have found a niche providing a service that attempts to address the issue of academic plagiarism in the digital age. This research project considers the effectiveness of Turnitin in identifying clearly copied written material from peer reviewed and published academic journal articles. It also considers both the effectiveness and the appropriateness of tasking plagiarism detection programs like Turnitin with the responsibility to identify plagiarism, particularly at a semantic level, in academic settings. In its consideration of future research directions, this project also offers reflections on the use of plagiarism detection programs and their place within a modern understanding of authorship and plagiarism in the digital age.

Finally, it is hoped that this project will contribute to a greater understanding of the efficacy of plagiarism detection programs and the appropriateness of their use in post-secondary educational settings, while providing new insights and research directions that will enhance our definitional understanding of plagiarism and authorship in the digital age.

2 Literature Review

This literature review examines the existing definitional understandings of plagiarism as it relates to written work for academic purposes, with a particular focus on post-secondary

writing. It also considers research surrounding the performance and pedagogical impact of web-based plagiarism detection programs in identifying plagiarism in written work from such perspectives as effectiveness, ease of use, deterrence, and ability to support student work around citation use and referencing.

2.1 Defining Plagiarism

The term “plagiarism” is most often defined as a series of behaviors and actions that range from the intentional or unintentional misuse of citations, such as paraphrasing, to the intentional copying of others’ works without proper attribution or credit. Park (2004) suggests that plagiarism has occurred when someone has represented a source’s “words or ideas as if they were one’s own, without crediting the source” (p. 292). Howard (1995) similarly defines plagiarism as “the representation of a source’s words or ideas as one’s own” (p. 799). Atkinson and Yeoh (2008) extend the definition to include the lack of a source’s consent or authorization, defining plagiarism as “the use of another’s work without proper acknowledgement or permission” (p. 222). Briggs (2009) suggests that plagiarism must have an element of intentional deception and “constitutes not simply the copying of someone’s work or ideas but rather the unacknowledged copying of such work and the subsequent submission of that work as one’s own” (p. 66).

2.2 Forms of Plagiarism

Using the above definitions as a guide, plagiarism can take several forms and be found across multiple academic disciplines, the details of which are examined below by several scholars.

Park (2004) describes a taxonomy of academic plagiarism by students that includes collusion, duplication, and commission:

- Collusion occurs when work that has been created by a group is represented as if it were the student's own.
- Commission is defined as the use of work by a student that is not his/her own and subsequently represented as if it were. Examples of commission might include the purchase of a paper from a commercial essay mill or Internet site or the submission of a paper written by another author as one's own.
- Duplication is defined as the reproduction of the same or almost identical work for more than one purpose, such as submitting the same assignment for multiple courses.

Briggs (2009) suggests that plagiarism can be subdivided into separate parts — “plagiarism of ideas, word-for-word plagiarism, and plagiarism of sources and authorship” (p. 66). Referring specifically to paraphrasing, Kakkonen and Mozgovoy (2010) suggest plagiarism often involves unsourced methods meant to avoid plagiarism detection, including the adding or removing of words or characters from written work; adding intentional and deliberate spelling and grammar mistakes; inserting words with the same or similar meanings; and the use of alternative syntactic forms for the same or similar expressions. Howard (1995) posits that plagiarism has occurred “when a writer fails to supply quotation marks for exact quotations; fails to cite the sources of his or her ideas; or adopts the phrasing of his or her sources, with changes in grammar or word choice” (p. 799). Howard further sub-divides plagiarism into three different forms: cheating, non-attribution of sources, and patchwriting:

- Cheating includes “borrowing, purchasing, or otherwise obtaining work composed by someone else and submitting it under one's own name” (p. 799).

- Non-attribution involves “writing one’s own paper but including passages copied exactly from the work of another (regardless of whether it is published or unpublished or whether it comes from a printed or electronic source) without providing (a) footnotes, endnotes, or parenthetical notes that cite the source and (b) quotation marks or block indentation to indicate precisely what has been copied from the source” (p. 799).
- Patchwriting includes “writing passages that are not copied exactly but have nevertheless been borrowed from another source, with some changes” (p. 799).

Shi (2012a) notes the difficulty post-secondary students, in particular, have in understanding appropriate paraphrasing in order to avoid plagiarism. In a study that explored whether English second language (ESL) students and professors across academic faculties shared similar views on the use of paraphrased, summarized, and translated texts, the author notes that skill sets in paraphrasing and summarizing are complex. She suggests they are dependent on “one’s knowledge of content, the disciplinary nature of citation practices, and the rhetorical purposes of using citations in a specific context of disciplinary writing” (p. 134).

Research surrounding plagiarism has also made note of the cultural distinctions surrounding its definitional understanding. Pennycook (1996) suggests that a culturally conditioned reality exists around constructed definitions of plagiarism, with the term interpreted differently across cultural lines. At the same time, he argues that Western notions of ownership, authorship, and intellectual property are in themselves distinct and historically specific and “need to be seen as a very particular cultural and historical development” (p. 221). For Pennycook, the post-enlightenment rise of individualism brought with it a more modern definition of literary

plagiarism, one that focused on the unattributed borrowing of ideas and language and its classification as an offense against the rights of individual property and copyright.

Shi (2012a) also makes note of cultural distinctions surrounding the understanding of plagiarism. The author draws attention to the disparate cultural interpretations of plagiarism employed by ESL students. She cites the example of Chinese students, whom, she claims, have traditionally reported less concern about the formal citation of sources in written work, while often regarding the Western notion of plagiarism as alien. The author does suggest, however, that Chinese students increasingly have adopted the Western definition of plagiarism and private textual ownership.

2.3 Effectiveness of Plagiarism Detection Software

The development of modern digital technologies, such as the Internet and its many adaptations, has provided writers with previously unseen levels of access to textual materials and resources and has influenced the growing perception that plagiarism is an increasing problem in education today (Dahl, 2007; Fiedler & Kaner, 2010; Howard, 2007). As a direct result, the use of plagiarism detection programs in post-secondary education has increased significantly over the past decade (Graham-Matheson & Starr, 2013). In response, some scholars have attempted to evaluate the effectiveness of these programs in performing the task of identifying clearly copied or similar textual material in academic writing. Thus far, results have demonstrated inconsistency in this area, with varying results across competing studies.

Fiedler and Kaner (2008), as part of a study on the effectiveness of the plagiarism detection programs Turnitin and MyDropBox, sourced computer science literature downloaded from a password-protected publisher's website and submitted the papers without modification to both programs. The authors then compared the scores calculated by each program in identifying

copied textual material within the submitted articles. Both Turnitin and MyDropBox calculate the number of textual matches between a submitted document and work indexed by each service as a percentage, which is then placed on a colour-coded scale that runs from low to high, progressing from blue through green, yellow, orange, and red. A submission identified by the program as having no matching textual material would receive a blue result, while a fully matching submission (100% matching textual material) would be flagged red. Using this colour-coded reference guide, a detection rate of less than 24% matching material would score blue (no matching material) or green (less than 24% matching material). For 10 of the 13 submissions,

Turnitin reported a similarity code of blue or green indicating there was little similarity in the experimental submissions to the works the detection services searched even though the experimental submissions were 100% plagiarized.

(Fiedler & Kaner, 2008, p. 185)

Similarly, eight of the 13 submissions processed through MyDropBox reported a similarity code of blue or green, with the authors again describing these submissions as containing little similar matching text. It should be noted at this point that neither Turnitin nor MyDropBox claim to provide adjudication as to the occurrence of plagiarism in written submissions. For example, Turnitin describes itself as a tool meant to assist users in identifying sources that contain text similar to submitted works, with decisions as to whether or not plagiarism has occurred left to users to decide (Turnitin Originality Check, 2016). Finally, while the programs examined in the study did not identify the complete original source of the majority of submitted articles, Turnitin did fully match three submissions, while MyDropBox matched four, to original primary sources in both public and journal publishing websites. As well, varying amounts of matching text were identified in the other experimental submissions.

Fiedler and Kaner conducted a second study in which 24 papers from a variety of peer reviewed education publications were downloaded from academic databases and submitted without modification to both Turnitin and MyDropBox. In this study, Turnitin “reported a green rating for 21 of the 24 experimental submissions,” defined as matching material between 1% and 24%, while “MyDropBox reported a green rating for 18 of 24 experimental submissions” (Fiedler & Kaner, 2010, p. 39). Turnitin did not identify the original primary source of any of the submitted articles, while MyDropBox identified two of the primary sources of the submissions. Varying amounts of matching textual material was found throughout all of the experimental submissions.

The authors concluded their second study by suggesting that because Turnitin and MyDropBox do not have access to certain bodies of professional literature, which explains their failure to identify significant instances of copied material, faculty and institutions addressing plagiarism detection would be wise to supplement their use with other investigative means. While they see some value in using plagiarism detection programs to expose obvious plagiarism quickly, they warn against placing too much trust in the software’s ability to consistently detect copied material in written work. They conclude with a recommendation that educators who use Turnitin and MyDropBox contact professional society executives and journal editors and demand that they enable efficient plagiarism checking of the articles in the respective databases.

In another study, Hill and Page (2009) selected 20 of their own personal undergraduate and graduate academic papers and submitted them to two plagiarism detection programs — Turnitin and SafeAssign. The submissions were separated into four groupings of five documents each. Plagiarized material was added to selected papers from a variety of sources used by both Turnitin and SafeAssign to check for originality. These sources included web pages, freely

accessible web-based databases such as Google Scholar and Wikipedia, and proprietary databases such as ProQuest. The first set contained five unaltered papers. Three more sets of papers were organized and altered with inserted textual material from a combination of the above sources. The additional content was copied and added directly without alteration. On average, submitted documents consisted of 15% plagiarized material and 85% original material. The findings of Hill and Page differed significantly from the findings of Fiedler and Kaner (2010). Hill and Page found Turnitin and MyDropBox to be significantly more successful at identifying clearly copied textual material in submitted works. Of the two examined programs, they found that Turnitin was the more accurate software with an 82.4% detection rate overall, compared with 61.9% for SafeAssign. According to the authors, both platforms correctly detected 70% or more of copied material added to the submissions from subscription databases and public web pages, while noting the programs had more difficulty detecting material plagiarized from freely available databases such as Wikipedia, FindArticles, PubMed, and Google Scholar. It is worth noting that Hill and Page used a markedly different methodology than Fiedler and Kaner (2010), submitting papers the authors had written as undergraduate and graduate students and adding clearly copied textual material to them before submitting them for an originality assessment. This was distinct from Fiedler and Kaner, who submitted verbatim-copied published academic articles for assessment. This difference in method, particularly the adding of copied text from multiple sources by Hill and Page, may have increased the likelihood of detection given the software's ability to prospect and match source material in multiple places. This was in contrast to Fiedler and Kaner, who sourced their submissions from selected password protected journal providers, some of which were not accessible by Turnitin and MyDropBox at the time of the study. Although Fiedler and Kaner detected matching textual material at a lower rate than Hill and

Page, both studies demonstrated the programs' inconsistency in identifying copied or similar written work in academic writing, with blind spots detected in both subscription and open content databases.

In a more recent study, Hunt and Tompkins (2014) compared the effectiveness of SafeAssign and Turnitin in detecting plagiarism in students' written submissions. The authors collected 293 samples of writing from first-year university students across several academic disciplines, including religion, psychology, and mathematics. Each sample was submitted to SafeAssign and Turnitin and analyzed for matching content. An analysis of the data revealed that 6.95% of all text submitted to SafeAssign was found to match existing material in the program's searchable databases, of which 4.26% was defined to be:

false positives, or matches between the student text and the database text that most likely did not represent instances of intentional plagiarism. (Hunt & Tompkins, 2014, p. 66)

This false positive matching was further defined as textual matches that included sourced quotations, references, and common language such as book titles, notable names, and common phrases. When the false positive content was removed, 2.75% of submitted textual content was found to be matching. A total of 7.64% of the text submitted to Turnitin was found to match existing material in the program's database, with 4.26% classified as false positive. When the false positive content was removed, 3.38% of submitted textual content was found to be matching. The authors concluded that there was no meaningful difference in the performance of the two programs, and suggested that both programs "grossly over-reported false positives and often failed to identify actual, blatant cases of plagiarism" (p. 69).

Other research examining the effectiveness of plagiarism detection programs has noted and defined the identification of certain matching textual material, such as sourced quotations and references, as well as random occurrences of text, and common language and phrasing, as false positives (Evans, 2006; Hill & Page, 2008; Kakkonen & Mozgovoy, 2010; Stapleton, 2012; Uzuner, Katz & Nahnsen, 2005). Evans (2006), referencing Turnitin specifically, says that because the program cannot distinguish plagiarized text from properly attributed content, it is not wholly reliable and should not be considered an accurate guide as to the amount of plagiarized text in a given submission. Hill and Page (2009) say the tendency of Turnitin and MyDropBox to produce false positives is concerning and call for users of these programs to examine flagged results closely and to do further investigation where necessary in order to promote more accurate findings. The authors urge that faculty and staff be trained specifically to interpret results in order to avoid negative assumptions and accusations of plagiarism.

Stapleton (2012), in an examination of both the effectiveness of Turnitin in identifying instances of copied textual material and as a deterrent to plagiarism, found repeated instances of what was described as “coincidental matching,” including shared quotations and references, and warned instructors employing Turnitin to take care in assessing results produced by the program (p. 131). Oghigian, Rayner and Chujo (2016), in a study examining the functionality and accuracy of Turnitin and its potential use in undergraduate science and engineering classes, report that the program routinely produced false positive results.

“While doing this analysis, it was immediately clear that the software produced a number of false positives” (Oghigian et al. p. 7). The authors identified false positive matches as shared references, quotations, tables, charts, and common expressions. Turnitin does not claim that results produced by the program are proof, or even an indication, of plagiarism. Nor does the

company claim that coincidental, or false positive matching, will not, or should not, occur in the provision of its service. Turnitin is clear in stating that indices provided by the program do not reflect an assessment of whether submissions have, or have not, been plagiarized. Instead, it claims its service is meant to serve as a tool to assist users in finding sources that contain similar text to submitted works. Any decision defining work as plagiarized is to be made by the person/instructor employing the program (Turnitin Originality Check, 2016). In fact, false positive matching may not be the most accurate term to describe these particular identified coincidental similarities in text, given Turnitin's stated function.

There has been some investigation, albeit preliminary, examining plagiarism detection at a semantic level, with an eye to improving the ability of plagiarism detection programs to detect more difficult cases of plagiarism, such as unsourced paraphrasing and the intentional plagiarizing of ideas. Maurer, Kappe and Zaka (2006), in a survey of research surrounding plagiarism, and methods to detect it in academic settings, considered the possibility of using plagiarism detection programs as a means of detecting document similarity using semantic analysis. The authors cite the work of Iyer and Singh (2005), who created a software program that extracts keywords (nouns, verbs, adjectives) from submitted documents and then runs comparison algorithms examining the syntactic structural characteristics of these submissions. If similarities are above a certain threshold, a more detailed sentence level examination is then conducted in an attempt to identify similar textual material. Iyer and Singh say this system has proven capable of detecting semantic/syntactic similarities in submitted works.

Uzuner et al. (2005) conducted another study measuring the effectiveness of using semantic analysis to identify document similarity. The authors used selections from novels translated into English as surrogates for actual plagiarized material. Many of the novels had been

translated on different occasions by several different translators over time. The authors then attempted to identify the copied material in the submissions using software designed to recognize syntactic and linguistic characteristics in sentences and key words. The results of the study suggested that identifying syntactic elements of expression that focus on changes in phrase structure may be an effective means of identifying more advanced and difficult forms of plagiarism. They concluded by suggesting that a consideration of linguistic information related to creative aspects of writing can improve identification of plagiarism by adding an important element to an evaluation of similarity (Uzuner et al. 2005).

Similarly, Gipp, Meuschke and Breitingner (2014) explored shortcomings in current automated plagiarism detection programs, namely their dependence on character-based verbatim matching, in a study examining the use of citation patterns as a means of identifying more difficult forms of plagiarism, such as intentionally deceptive and unsourced paraphrasing, and the plagiarism of ideas. The authors examined a collection of peer reviewed biomedical texts as part of their investigation, applying designed algorithms to their test data set and making use of the semantic information implied by the citations found within sampled texts. The study identified and analyzed similar patterns in the citation sequences of tested documents and noted similarities. The authors claim this allowed for the detection of plagiarism that could not otherwise have been detected automatically by traditional text-based approaches. They say this citation-based detection approach significantly outperformed character-based approaches in identifying documents that contained paraphrased and structural idea similarity. In their findings, the authors also claim to have discovered “several cases of previously unidentified plagiarism” in the collection samples (Gipp et al. 2014, p. 1,540).

2.4 Plagiarism Detection Software: Pedagogical Application

Other research into plagiarism detection programs has included examinations of its deterrent effect on academic plagiarism, its overall ease of implementation and use, as well as its ability to support student learning around established writing conventions, including sourcing and referencing, particularly at the post-secondary level (Dahl, 2007; Badge, 2007; Batane, 2010; Evans, 2006; Lofstrom & Kupila, 2013; Rogers, 2009). This research provides further insight into the effectiveness and appropriateness of employing plagiarism detection programs as a means of uncovering instances of plagiarism in academic settings. Several studies (Buckley & Cowap, 2013; Dahl, 2007; Evans, 2006) have investigated the ease by which plagiarism detection software can be implemented and used in post-secondary classroom settings. While the focus and methodology of these studies differed to some extent, each contained surveys of post-secondary undergraduate and graduate students in order to gather feedback from students who used the programs. These studies reported that students and teachers had little trouble using plagiarism detection programs as a means of submitting and evaluating written work for plagiarism.

Sutherland-Smith and Carr (2005) found teachers to be “reasonably happy with the usability of the software” (Turnitin), stating that “instructions for use were clear and uploading documents to the Turnitin site was quite manageable” (p. 98). Dahl (2007) found that of 23 graduate students surveyed, 22 “strongly agreed” or “agreed,” in a five-point Likert survey, that Turnitin was “easy to use” (p. 189). In an exploratory study conducted by Evans (2006), undergraduate students were trained in the use of Turnitin and asked to submit term assignments using the program. They then were provided with a questionnaire to gather feedback on their user experience. Students’ assessment of the program and its ease of use were generally positive

(p. 92). However, Thompsett and Ahluwalia (2010) found more mixed results in a survey of undergraduate science students who used Turnitin as a tool to help improve citation and referencing skills. Students were provided with a research questionnaire that examined their overall experience using the program, with one section of the questionnaire focused on ease of use. The study found a majority of students surveyed (59%) did not find Turnitin easy to use, nor did they regard the program as a useful learning tool for improving their referencing skill sets. The authors suggest that a lack of appropriate prior training in the use of the program may have contributed to the negative responses.

Overall, research results have been mixed concerning the ability of plagiarism detection programs to support student learning and provide effective feedback about the proper use of citations, references, and writing conventions. According to Lofstrom and Kupila (2013), a significant cohort of students (41 of 53) felt the use of Turnitin as a primary means of submitting written work did not guide or positively impact their writing and sourcing. In an exploratory study, students' main criticism of the program and its ability (or lack thereof) to provide quality feedback on writing, sourcing, and referencing centered on the vagueness of the originality report provided by Turnitin, which calculates matching textual content as a percentage. Several students did, however, recognize the program's potential to support their learning and help them understand academic writing in a more profound way. This included assistance in how to make proper references to the work of others; how to re-evaluate their own practices in response to information provided by the program; and help in developing voice, style and expression in their written work.

Similarly, Rogers (2009) suggests that secondary school students also had mixed opinions surrounding the feedback offered by Turnitin and the program's ability to assist them to

become better writers and to be more aware of established conventions around sourcing and referencing. The author reports that a significant number of students felt the service did not help them become better writers. In terms of deterrence, research suggests that the use of plagiarism detection programs can lead to a decrease in instances of plagiarism over time at the post-secondary level, with students less likely to knowingly plagiarize if they are aware in advance that assignments will be screened with detection programs (Badge, Cann & Scott; 2007; Batane 2010). Batane (2010) measured plagiarism levels in written assignments submitted by undergraduate university students across multiple academic faculties at the University of Botswana before and after the implementation of Turnitin as a method of screening written works, and found a deterrent effect of 4.3% relating to cases of plagiarism identified by the program. Badge et al. (2007) examined the deterrent impact of JISC Plagiarism Detection Service at the University of Leicester, and found that while detected instances of plagiarism increased across the institution during the first year of implementation, plagiarism declined in subsequent years. The authors concluded by stating the “results suggest that the JISC PDS is an effective means of both detecting and deterring plagiarism” (Badge et al., 2007, p. 437).

2.5 Summary of the Literature Review

In summary, the detection rates of several plagiarism detection programs have varied significantly in published studies, from less than 20% (Fiedler & Kaner, 2008; Fiedler & Kaner, 2010), to more than 80% (Hill & Page, 2009). While this research is useful in providing insight into the effectiveness of specific plagiarism detection programs and their use in academic settings, there would appear to be room to expand on these studies. This could include an examination of other plagiarism detection software programs, while also building larger and more representative measurements, with a goal of producing more accurate results.

Second, research examining the effectiveness of plagiarism detection programs has noted a tendency of these programs to identify common matching textual material, such as sourced quotations, references, and common language and phrasing, in the production of user originality reports (Evans, 2006; Hill & Page, 2008; Kakkonen & Mozgovoy, 2010; Stapleton, 2012; Uzuner et al. 2005).

To this point, there has been limited research on the effectiveness of plagiarism detection programs in identifying plagiarism at a semantic level. The research that has been completed has been largely theoretical (Gipp et al. 2014; Iyer & Singh, 2005; Maurer et al. 2006; Uzuner et al. 2005). There would seem to be room to build on these early investigations, particularly around the effectiveness and appropriateness of tasking plagiarism detection software with identifying academic plagiarism at a semantic level. New research might also include an examination of whether the ability to investigate the plagiarism of ideas could be built into existing commercial plagiarism offerings.

Thus far, research suggests that plagiarism detection software programs such as Turnitin, SafeAssign, MyDropBox, and JISC have some deterrent effect on student plagiarism in post-secondary classrooms, particularly when students are informed in advance that these programs will be used to screen written work. Similarly, several studies have demonstrated that both students and teachers have little trouble employing plagiarism detection programs in post-secondary classrooms as a means of submitting written work, particularly if they are instructed in how to use the programs in advance. In terms of the usefulness of plagiarism detection software as a learning support around the use of writing, sourcing, and referencing conventions in written academic work, results have thus far been mixed among post-secondary and secondary students.

Questions remain regarding the ability of plagiarism detection programs to provide effective feedback in these areas.

3 Methods

I conducted an exploratory study for the purpose of assessing the effectiveness of the plagiarism detection program Turnitin in identifying copied textual material in written sample submissions. This study intended to answer three research questions:

3.1 Research Questions

1. Is Turnitin adequate in identifying instances of textual similarity in copied written submissions?
2. Is Turnitin adequate in identifying primary sources of text in copied written submissions?
3. Is Turnitin able to detect plagiarized ideas and content in written submissions?

3.2 An Exploratory Study Design and Procedure Using Turnitin

For the study, I selected 10 peer reviewed, published academic articles from a variety of academic disciplines including humanities, social sciences, education, science, and engineering. The search term “plagiarism detection” was used and processed through the main search engine at the University of Ontario Institute of Technology (UOIT) library. Sample submissions were selected from peer reviewed journals and downloaded from online academic databases sourced by the UOIT library. This collection of databases covers content across an expansive range of academic disciplines including the humanities, mathematics, biology, chemistry, physics, material science, health science, social science, business and engineering. It also provides a range of content materials including journals, theses, conference papers, eBooks and streaming videos.

The sample articles ranged in publication date from 1970 to 2013 (see Appendix A for detailed descriptions of the 10 articles). Each article was downloaded in PDF form and then copied verbatim into a Word document and submitted to Turnitin for an originality check without modification. Published articles from recognized academic journals were chosen as the source of these submissions because samples of this kind would represent blatantly plagiarized academic material. I believed that these sample submissions likely would be located in searchable online journal and library databases, as well as in other online sources of content, and would provide a fair measure of Turnitin's overall detection ability, particularly in the areas of text matching and primary source identification. Similarly, due to privacy restrictions surrounding student work processed through Turnitin, I believed that sample submissions taken from published and publicly available journal offerings would produce findings that were more accessible for the present research purposes because they would not be subject to the privacy protection provided for student works. The evaluation of the program included an assessment of whether Turnitin was adequate in identifying instances of textual similarity in copied submissions; whether the program could identify the primary source origins of submissions; and whether the program was able to detect plagiarized ideas and content.

When work is submitted to Turnitin, it is compared against content databases using string-matching algorithms that enable the direct comparison of submitted textual material with both current and archived online sources (Kakkonen & Mozgovoy, 2010). The program also compares submissions to student papers that have been submitted previously by users of Turnitin and archived for future reference. Turnitin compares submitted textual material to a variety of databases provided by its content partners, including textbook publishers, digital reference

collections, library databases, subscription-based publications, homework helper sites, and books (Turnitin Content, 2016).

Turnitin software then processes the sample submissions and the program generates a similarity report for each. The similarity report calculates the percentage of copied or similar content in the provided submissions against content matched to other sources, along with information indicating the origins of copied textual material. Turnitin reports on similarities between submitted textual documents and written works. These include published academic articles, student essays, and textual content from a variety of online sources, such as websites and databases, indexed by the program in its text-matching searches. The program produces a colour-coded similarity index with a percentage score indicating the amount of matching text uncovered (Turnitin-Guide, 2016). Scores can range from 0 to 100%. A blue indication suggests that no matching or similar text was found. A green indication represents matching or similar text found between 1% and 24%. A yellow indication represents 25% to 49% matching material. An orange indication represents 50% to 74%, and finally, a red indication represents matching or similar textual content of 75% or above contained in the sample document.

Similarity Index

The color of the report icon is linked to one of five tiers of the Similarity Index. This is based on the amount of matching text found by the repository comparison. The possible similarity index percentage ranges are linked to a corresponding color:

- **blue** (no matching words)
- **green** (one matching word - 24% similarity index)
- **yellow** (25-49% similarity index)
- **orange** (50-74% similarity index)
- **red** (75-100% similarity index)

The overall similarity index and the corresponding color scaling of the icons provides a basic indication of how much information contained in a specific submission is matched to other sources in the Turnitin repository. This number is a raw amount of matching done against the repositories selected for the assignment the submission was made to. Direct quotation, citations, or bibliography areas of the paper are not automatically excluded. The decision to permanently exclude or disregard matches to these types of text in a paper is made solely by the instructor of the class.

Figure 1. Turnitin Similarity Index (Turnitin Originality Check, 2016).

These indices do not directly represent an assessment from Turnitin regarding whether or not submitted content should be considered plagiarized. A final adjudication and determination of plagiarism is beyond the program's current capabilities, as is indicated in Turnitin's recommendations to users attempting to interpret originality findings (see Figure 2). Users employing the program are encouraged to use results identifying instances of matching or similar textual material as part of a larger process of determining if these textual matches are acceptable or not, as opposed to considering them as a final arbiter of whether or not plagiarism has occurred (Turnitin Originality Check, 2016).

Warning: These indices do not reflect Turnitin's assessment of whether a paper has or has not been plagiarized. Originality Reports are simply a tool to help an instructor find sources that contain text similar to submitted papers. The decision to deem any work plagiarized must be made carefully, and only after in depth examination of both the submitted paper and suspect sources in accordance with the standards of the class and institution where the paper was submitted.

Figure 2. Turnitin Assessment Guidelines (Turnitin Originality Check, 2016).

The originality check produced by Turnitin for each sample submission was documented and recorded in a table (see Appendix A). Sample articles were itemized under several categories that were created as a way of organizing evidence related directly to the research questions. These included the title of the sample submission, the year of publication, the author(s), a description of the percentage of matching textual content, a description of whether the original primary source was identified by Turnitin, and a comment section analyzing overall findings. Observational findings were catalogued and described in the comments section and contained a detailed analysis of the effectiveness of Turnitin in identifying matching text in each sample submission, as well as whether or not Turnitin was able to identify the primary source of sample submissions.

4 Results

This exploratory research offers insight into the degree to which Turnitin is adequate in identifying clear examples of textual similarity in sample submissions; whether the program successfully identifies the primary source origins of these submissions; and finally whether the program is able to detect plagiarized ideas and content in written submissions. In the following subsections, I present the results in relation to the research questions posed in section 3.1.

4.1 Success in Identifying Textual Similarities and Primary Sources of Copied Written Submission

Research Questions 1 and 2 from the exploratory study sought to determine the degree to which Turnitin successfully identified instances of textual similarity in copied written submissions as well as the primary sources of those copied submissions. The results indicated that Turnitin identified some matching text in all ten submitted works ranging between 6% and 100% (see Appendix A). Turnitin also successfully identified a majority of original primary matching content in submissions 4 through 10, which also were the seven most recent submissions, published between 1986 and 2013 (see Appendix A). These articles received similarity scores between 93% and 100%, with matching content located by Turnitin primarily in established online academic publishing databases and in the program's own repository of submitted student work. Small examples of matching content were also sourced to several Internet sources.

In submission four, the majority of textual content (99%) matched the original primary source published in Taylor and Francis Online, a subscription-based publisher of academic journals and books (Taylor and Francis, 2016). Some content (4%) matched several Internet sites and student papers sourced by Turnitin. This consisted mostly of shared references and minor

textual similarities in select sentences. Unmatched textual material (1%) consisted primarily of random selections of matching text from the original submission.

In submission five, the majority of textual content (95%) matched the original primary source published in Gale Publishing, a subscription-based online publisher of academic journals and books (Gale Publishing, 2016). Similarly, a partial preview of the original article submission was found in Questia, an online research and essay writing service (Questia, 2015). Unmatched textual content included search terminology and random text samplings from the original submission.

In submission six, all textual content (100%) matched the original primary source in Gale Publishing and the Ontario Ministry of Education Ontario Software Acquisition Program Advisory Committee website — osapac.ca. A partial preview of the article submission was also matched to Questia, while small sections of content, including partial sentences and shared references, were found on several Internet sites and student papers sourced by Turnitin.

In submission seven, the majority of textual content (99%) matched the original primary source in two academic library databases — the University of Florida and McMaster University — and in ScienceDirect, an academic database of scientific journal articles (ScienceDirect, 2015). Small examples of matching textual content (under 3%) were also found in student papers sourced by Turnitin, including shared references and parts of sentences. Unmatched textual content appeared to consist of random occurrences of similar text.

In submission eight, all textual content (100%) matched the original primary source in the Canadian Centre of Science and Education database. This is a not-for-profit education resource provider that publishes scholarly journals in a wide range of academic fields, including social sciences, humanities, education, economics, and natural and medical sciences (Canadian Centre

of Science and Education, 2015). Selections of textual content (8%) were found in several student papers sourced by Turnitin. Matches consisted of similarities in written content, including partial sentences and shared references. Small content matches (4%) were also found in several academic articles sourced by Turnitin in a number of online publishers, including shared references and partial sentences.

In submission nine, the majority of textual content (93%) matched a copy of the original PDF version of the article published on the author's personal website. A single reference was also sourced to IEEEExplore Digital Library, a technology-based professional association that offers online academic journal holdings for members (IEEEExplore Digital Library, 2015). Unmatched textual material consisted of random occurrences of text.

In submission ten, the majority of textual content (99%) matched the original primary source in Springer Publishing, a publisher of science-based books and journals (Springer Publishing, 2015). Ten percent of the original article content matched student papers sourced by Turnitin, including shared references and sentence parts. Small selections of textual content (1%) also matched a variety of Internet sources. Unmatched textual content appeared to consist of random occurrences of similar text.

To summarize, in submissions 4 through 10, each of which received similarity scores ranging from 93% to 100%, the majority of original or primary textual content was sourced to recognized online academic and/or professional journal publications, or in one case, the author's own personal website. Small selections of content (10% or less) matched other sources, including Internet sites and student papers sourced from a number of academic institutions that use Turnitin as a resource for plagiarism detection. These similarities were largely attributed to commonalities in academic sourcing, including the use of similar references. Partial sentence

matches were also found, however this matching material typically consisted of small sections of text (typically one sentence or less) and in no cases was there evidence of substantial copying of original content.

However, a gap was noted in the identification of the original primary source in three of the ten submissions. Articles 1, 2, and 3 (see Appendix A), whose publication dates were 1970, 1976, and 1982, had similarity ratings of 16%, 6% and 7% respectively. In each case, Turnitin failed to identify the original primary source of the submission in published online academic and/or professional journal publications, although each was originally sourced through the University of Ontario Institute of Technology's student library reference search portal and came from recognized academic journals.

In submission one, the majority of textual content (84%) was not matched by Turnitin. Small selections of textual content, including references and minor examples of matching text, were found in several academic articles sourced to publishers including Wiley Online Library (Wiley, 2016), Taylor and Francis Online (Taylor and Francis, 2016), and Sage Journals (Sage Journals, 2016). The remainder of matching textual material was found in student papers sourced by Turnitin. This included shared references and parts of sentences.

In submission two, the majority of textual content (94%) was not matched by Turnitin. Some content (3%) was found on a now-inactive Internet site (scholarlywriting.net), including several partial sentences and a shared quotation. Some shared references and full and partial sentences (3%) were sourced to student essays.

In submission three, the majority of textual content (93%) was not matched by Turnitin. Three percent of submitted matching textual material was found in several student papers catalogued by Turnitin. Several shared references were matched to a single academic article

sourced through Taylor and Francis Online. The remainder of matching textual content included shared references, quotations, and sentence fragments from several Internet sites.

4.2 Detecting Plagiarized Ideas and Unsourced Paraphrased Content

To answer research question 3, Turnitin's ability to identify the presence of plagiarized ideas and paraphrased yet unsourced content was limited. It was unable to indicate whether textual submissions were similar from a semantic perspective. Identifying more advanced and difficult forms of plagiarism were beyond the software's capability and scope of design. In the examination of the 10 sample submissions processed for the present study, no evidence was found demonstrating an ability on the part of Turnitin to detect semantic matches in textual content. And while Turnitin did identify and report on a wealth of matching textual content, this identification was a result of the program's algorithmic text matching function. Simply put, text had to be matched identically to be identified by the program. Turnitin did not offer or provide any other mechanisms for identifying clearly plagiarized material, such as a semantic evaluation of suspect text, namely the intentional plagiarism of ideas.

5 Discussion

Findings from the exploratory study identified several areas for discussion surrounding both the use and effectiveness of plagiarism detection programs such as Turnitin. These findings are discussed in detail in the following section, including noted gaps in the ability of Turnitin to identify matching content in password protected databases; the impact of submission age on detection; the identification of coincidental matching textual material in program results, and difficulty in identifying semantic matches.

5.1 Gaps in Identifying Textual Similarity and Primary Sources

This exploratory research identified several gaps in Turnitin's ability to identify instances of textual similarity and primary sources of text in written submissions. The program did not identify the original primary textual source in three of ten submissions. Of the seven articles in which the majority of original primary content was identified as matching (see Appendix A), six were matched directly to subscription-based publisher databases and/or university library databases, while one was matched to the original author's own personal website. These results suggest that the program's publisher content partnerships are the key to enabling access to published academic material, particularly original primary content that is password protected in publisher or library databases.

A number of potential issues might account for Turnitin's difficulty in detecting the content source of the three examples mentioned. First, the age of the submitted publications may have influenced their likelihood of detection. For example, article submissions, 1, 2, and 3, were the oldest of the sampled articles by publication date (published in 1970, 1976, and 1982 respectively) (see Appendix A). Older journal articles may, in fact, be less likely to be included in searchable online databases and thus more difficult for plagiarism detection programs to detect. Several academic studies (Fiedler & Kaner, 2008; Fiedler & Kaner, 2010; Hill & Page, 2009; Kakkonen & Mozgovoy, 2010; Maurer et al., 2006) previously identified inconsistencies in the ability of plagiarism detection programs to identify clearly copied textual material in both subscription-based and open online databases, although none identified submission age as a possible cause. This finding could also be related to the possibility that articles with older publication dates may be less likely to be sourced digitally in student papers and thus more difficult for plagiarism detection programs to detect in digital searches. Even when content

matches were made to student papers found in Turnitin's archive, little detail or context surrounding the matching material could be accessed. Archived student papers are protected by privacy agreements between the program provider and member institutions, with access to student submissions limited to the participating student and supervising instructor. While matching content is flagged and noted, users cannot access student papers that contain matching material connected to the submitted work, thus making it more difficult to determine whether plagiarism has occurred. Instructors attempting to examine the origins of matching content in an investigation of plagiarism, for example, would find themselves limited in this regard. Similarly, authors of published works employing plagiarism detection services as a reverse plagiarism check would also find themselves limited in their ability to assess whether their work has been sourced properly once it is identified as matching in student works. This is largely due to the lack of access provided to student content. Student papers are the only content afforded this level of privacy by Turnitin.

Turnitin has established content partnerships with a large number of providers, including text-book publishers, digital reference collections, library databases, and subscription-based publications (Turnitin Content, 2016). However, Turnitin lacks agreements with some subscription-based content providers, thus creating gaps in the program's ability to access and identify published content online. For example, the three articles identified in the exploratory study, in which Turnitin was unable to match the majority of original primary content, were originally sourced from JSTOR, a not-for-profit shared digital library service with a focus on the digitization and cataloguing of scholarly content (JSTOR, 2016). JSTOR has content agreements with several academic libraries and publishers but currently does not have an agreement with iParidigm, the parent company of Turnitin. Therefore, Turnitin does not have access to JSTOR's

password-protected content database. While Turnitin did match small selections of content in articles 1, 2, and 3, including shared references and small samples of matching text to other sources (see Appendix A), the majority of primary textual content was not detected by the program in its search.

This gap, created by a lack of access to content databases, was also noted by Fiedler and Kaner (2008) in their examination of Turnitin and MyDropBox. The authors found that a majority of submitted software engineering articles sourced from the publisher IEEE Xplore Digital Library, and submitted to Turnitin and MyDropBox as if they were the authors' own, were not detected by either program. The authors traced this vulnerability in the two programs' ability to detect the original primary source of submitted articles to their lack of access to the IEEE Xplore password-protected database. The few submissions that were identified were traced to copies of articles that had been posted to the public web, as opposed to the original password protected site. They concluded their study by recommending that journal providers and plagiarism detection program providers work towards improving content access in the future:

The professional societies must work out a licensing structure that gives plagiarism detection services access to professional literature so that teachers, editors and manuscript reviewers, can time-efficiently determine whether submitted work has been plagiarized.

(Fiedler & Kaner, 2008, p. 187)

Hill and Page (2009), in their study examining the efficacy of Turnitin and SafeAssign, found in some circumstances the programs failed to detect textual material clearly copied from both subscription-based and open databases. However, unlike Fiedler and Kaner (2008), and the preliminary study conducted for this paper, their findings did not identify a vulnerability related to a specific journal provider. Similarly, Hunt and Tompkins (2014) compared the effectiveness

of SafeAssign and Turnitin in detecting plagiarism in written work collected from undergraduate students across a variety of academic disciplines. The authors found that both programs at times failed to detect matching textual material in both subscription-based and open databases, although they also did not identify a vulnerability related to a specific journal provider.

The exploratory study conducted for this research project, as well as several previous academic studies (Evans, 2006; Hill & Page, 2008; Hunt & Tompkins, 2014; Oghigian et al. 2016; Uzuner et al. 2005), identified false positive or coincidental matching as a potential concern for those employing plagiarism detection programs as a means of identifying plagiarism in academic work. Findings revealed that programs such as Turnitin do not distinguish between properly identified and sourced quotations and references, as well as random occurrences of text, and routinely represent these findings in similarity indexes, expressed as a percentage of copied material in submitted work. This reality was described by Uzuner et al. (2005):

Using keyword overlaps to identify plagiarism can result in many false negatives and positives: . . . overlap in ambiguous keywords can falsely inflate the similarity of works that are in fact different in content. (p. 37)

Hunt and Tompkins (2014) say that SafeAssign and Turnitin “grossly over-represented false positives” in their assessment of written student submissions (p. 69). Oghigian et al. (2016), found Turnitin produced a number of false positives, including matching references, quotations, tables, charts, and common expressions. Stapleton (2012), found similar results and warned instructors to take care in assessing results produced by the program. Kakkonen and Mozgovoy (2010) suggest that given the volume of textual material available online, it is very possible for student work to coincidentally resemble existing academic work. Similarly, it is also possible for plagiarism detection programs, as currently constructed, to identify content, such as shared

references, shared quotations, or parts of sentences that are matching but not plagiarized. Hill and Page (2009) say that the rate of false positive detection produced by plagiarism detection programs reinforces the need for users to examine flagged results closely and effectively for accurate and correct detection. They suggest faculty and staff be trained specifically in interpreting results so as to avoid negative assumptions and accusations of plagiarism.

Finally, while Turnitin often produces coincidental matching in the provision of its originality indices, this is a result of the stated goal or function of the program, which is essentially the matching of textual content from submitted sources to existing sources found online. Turnitin does not claim that coincidental, or false positive matching, will not, or should not, occur in the provision of its text matching service. In fact, it states clearly that indices provided by the program do not reflect an assessment and judgment of whether submissions have, or have not, been plagiarized. Instead, Turnitin explicitly indicates that results are meant to provide a tool aimed at assisting users find sources that contain text similar to submitted works. Final decisions regarding whether plagiarism has occurred should be made by those using the program and as part of a larger investigation of suspect work.

6 Conclusion

Academic literature, including this exploratory study, has demonstrated that plagiarism detection programs are inconsistent in identifying similar or copied content in submitted academic work. Specifically, the present study demonstrated a gap in Turnitin's ability to identify copied textual content in submissions originating from subscription-based repositories of academic work that did not have a content agreement with the program provider. This had a clear impact on the program's effectiveness and it would behoove plagiarism detection providers to shore up these content gaps wherever possible to ensure that they have access to necessary

content sources and thus improve the likelihood of detecting instances of plagiarism in submitted works.

Similarly, the study results suggest that the age of submitted publications may have a role in their likelihood of detection by plagiarism detection programs such as Turnitin. For example, article submissions 1, 2, and 3, were the oldest of the sampled articles by publication date (published in 1970, 1976, and 1982 respectively) (see Appendix A) and were not detected by Turnitin in its search. Older journal articles may, in fact, be less likely to be included in searchable online databases, including public and password protected databases, and thus more difficult for plagiarism detection programs to detect. Similarly, older articles may also be less likely to have been sourced in student papers archived by these programs.

The identification of false positives, or coincidentally matching textual material, was also noted as an area of concern about plagiarism detection programs and their effectiveness as a tool for identifying plagiarism in written works. Specifically, it was revealed that programs such as Turnitin do not distinguish between properly sourced material such as quotations, references, and random occurrences of text, routinely representing these findings in their similarity indexes. Moving forward, it is important that users of these programs examine matching results closely and effectively for accurate and correct detection so as to avoid negative assumptions and accusations.

Finally, as currently designed, plagiarism detection programs cannot test for plagiarism at a semantic level, as identification, assessment and interpretation of more advanced plagiarism methods, such as the intentional plagiarism of ideas or purposed unsourced paraphrasing, fall outside the purview of their design and function. That said, despite these limitations, web-based plagiarism detection programs can still provide educators with much needed assistance in

detecting plagiarism in student work, particularly in cases of verbatim copying. These programs offer efficiencies of use, for example the ability to conduct multiple searches at once, and can provide an effective starting point for investigations of suspected plagiarism cases, while serving as part of a multifaceted approach in dealing with academic plagiarism in the digital age.

7 Future Research

To conclude, the use of plagiarism detection software and its place in gaining an understanding of plagiarism and authorship in the digital age need to be discussed with an eye to future research. To this point, much of the academic literature surrounding plagiarism detection programs has focused on examining their effectiveness at identifying copied textual material (Hill & Page, 2009; Kaner & Fiedler, 2010; Hunt & Tompkins, 2014), their deterrent effect (Badge, 2007; Batane, 2010), their ability to provide feedback around citation and sourcing practices (Lofstrom & Kupila, 2013; Rogers, 2009), and their ease of implementation and use in academic settings (Buckley & Cowap, 2013; Dahl, 2007; Evans, 2006). Rhetorical arguments have also attempted to establish an improved definitional understanding of plagiarism and authorship in the digital age, while at the same time considering the propriety of using plagiarism detection software in academic settings (Fitzpatrick, 2011; Howard, 2007; Jenson & de Castell, 2004; Reymen, 2010). Moving forward, there would seem to be a place for further investigation on all of these fronts.

Further examination of plagiarism detection software programs, and specifically their ability to identify plagiarized academic material, is needed. This includes a particular focus on limitations surrounding their access to content providers, as well as in detecting more advanced forms of plagiarism. Similarly, several studies have examined the efficacy of plagiarism detection programs, including their deterrent effect, their ability to provide feedback around

citation and sourcing practices, as well as their ease of use and implementation in classroom settings. However, these examinations have for the most part focused on the post-secondary level. As these programs increase their presence in secondary classrooms, an examination of their impact on that level of education would seem appropriate.

There would also appear to be a need to continue to examine the appropriateness of using plagiarism detection programs in today's evolving digital age, particularly as the definitional understanding of authorship and originality continues to evolve. The development of modern technologies, such as the Internet and its endless applications, have in some ways helped to further entrench an individual and commercial definition of authorship that is served by plagiarism detection programs. However, these same applications may also, in their own way, be beginning to confront this reality. Digital technologies now provide writers with unending access to textual resources and are precipitating a shift in our understanding of these terms. As Woodmansee (1994) suggests, the computer is "dissolving the boundaries essential to the survival of our modern fiction of the author as the sole creator of unique and original works" (p. 25). And while it is true that the perception of plagiarism as a rising tide in education has led to an increasing reliance on the use of plagiarism detection programs, such as Turnitin, in academic settings (Graham-Matheson & Starr, 2013), at the same time, digital technology and its associated applications have introduced many tools that provide new ways to manipulate, appropriate, write, rewrite, and link to written materials. In many respects this has served to mitigate or relinquish the illusion of creating autonomous textual works and directly confronts modern definitions of authorship and originality (Jenson & de Castell, 2004). It is worth considering how the use of plagiarism detection software fits with evolving methods and

definitional understandings of textual construction, authorship, and originality in modern digital environments.

References

- Anderson, D., & Wheeler, D. (2010). Dealing with plagiarism in a complex information society. *Education, Business and Society: Contemporary Middle Eastern Issues*, 3(3), 66-77. doi: 10.1108/17537981011070082
- Atkinson, D., & Yeoh, S. (2008). Student and staff perception of the effectiveness of plagiarism detection software. *Australasian Journal of Educational Technology*, 24(2), 222-240. doi 10.1.1.115.3076
- Austin, J., Brown, L. (1999). Internet plagiarism: Developing strategies to curb student academic dishonesty. *The Internet and Higher Education* 2(1), 21-33. doi: 10.1016/S1096-7516(99)00004-4
- Badge, J., L, Cann. A. J., & Scott, S. (2007). To cheat or not to cheat? A trial of the JISC plagiarism detection service with biological sciences students. *Assessment and Evaluation in Higher Education*, 32(4), 433-439. doi: 10.1080/02602930600898569
- Batane, T. (2010). Turning to Turnitin to fight plagiarism among university students. *Education Technology & Society*, 13(2), 1-12. Retrieved from <http://search.proquest.com.uproxy.library.dcuoit.ca/docview/1287029764?accountid=14694>
- Braumoeller, B.F., & Gaines, B.J. (2001). Actions do speak louder than words: Deterring plagiarism with the use of plagiarism-detection software. *Political Science and Politics*, 34(4), 835-839. Retrieved from <http://www.jstor.org.uproxy.library.dc-uoit.ca/stable/1350278>
- Briggs. R. (2009). Shameless! *Journal of Theoretical Humanities*, 14(1), 65-75. doi: 10.1080/09697250903006476

- Brinkman, Bo. (2013). An analysis of student privacy rights in the use of plagiarism detection systems. *Science Engineering Ethics*, 19(3), 1255-1266. doi: 10.1007/s11948-012-9370-y
- Buckley, E., & Cowap, L. (2013). An evaluation of the use of Turnitin for electronic submission and marking and as a formative feedback tool from an educator's perspective. *British Journal of Educational Technology*, 44(4), 562-570. doi: 10.1111/bjet.12054
- Canadian Centre of Science and Education. (2016, March 23). About us. Retrieved from <http://web.ccsenet.org/about-us.html>
- Carroll, J. (1982). Plagiarism: the unfun game. *The English Journal*, 75(5), 92-94. Retrieved from <http://www.jstor.org/stable/816218>
- Chandrasoma, R., Thompson, C., & Pennycook, A. (2004). Beyond plagiarism: Transgressive and non-transgressive intertextuality. *Journal of Language, Identity & Education*, 3(3), 171-193. doi: 10.1207/s15327701jlie0303_1
- Curwood, J.S., Fields, D., Lammers, J., & Magnificio, A. M. (2014). DIY media creation. *Journal of Adolescent and Adult Learning*, 58(1), 19-24. doi: 10.1002/jaal.331
- Dahl, S. (2007). Turnitin: The student perspective on using plagiarism detection software. *Active Learning in Higher Education*, 8(2), 173-191. doi: 10.1177/1469787407074110
- DeVoss, D., & Rosati, A. (2002). 'It wasn't me, was it?': Plagiarism and the web. *Computers and Composition*, 19(2), 191-203. doi: 10.1016/S8755-4615(02)00112-3
- Elfadil-Eisa, T.A., & Salim, N. (2015). Existing plagiarism detection techniques. *Online Information Review* 39(3), 383-400. doi: 10.1108/OIR-12-2014-031
- Endeshaw, A. (2005). Reconfiguring intellectual property for the Information Age. *The Journal of World Intellectual Property*, 7(3), 327-364. doi: 10.1080/13600830500376873

- Ercegovic, Z., & Richardson, J. (2004). Academic dishonesty, plagiarism included, in the digital age: A literature review. *College & Resource Libraries*, 65(4), 301-318. doi: 10.5860/crl.65.4.301
- Evans, R. (2006). Evaluating an electronic plagiarism detection service: The importance of trust and the difficulty of proving students don't cheat. *Active Learning in Higher Education*, 7(1), 87-99. doi: 10.1177/ 1469787406061150
- Fiedler, R. L., & Kaner, C. (2008). A cautionary note on checking software engineering papers for plagiarism. *IEEE Transactions on Education* 51(2), 184-188. doi: 10.1109/TE.2007.909351
- Fiedler, R. L., & Kaner, C. (2010). Plagiarism detection services: How well do they actually perform? *IEEE Technology and Society Magazine*, 29(4), 37-43. doi: 10.1109/MTS.2010.939225
- Fitzpatrick, K. (2011). The digital future of authorship: Rethinking originality. *Culture Machine* 12(1), 2-26. Retrieved from <http://www.culturemachine.net/index.php/cm/article/view/download/interstitial/433/466>
- Flint, A., Clegg, S. & Macdonald, R. (2006). Exploring staff perceptions of student plagiarism. *Journal of Further and Higher Education*, 30(2), 145-156. doi: 10.1080/03098770600617562
- Fox, M. (1994). Scientific misconduct and editorial and peer review processes. *The Journal of Higher Education*, 65(3), 298-309. Retrieved from <http://www.jstor.org/stable/2943969>
- Gale Cengage Learning. (2016, March 21) About us. Retrieved from <http://solutions.cengage.com/gale/about/>

- Gibelman, M., Sheldon, R., & Fast, J. (1999). The downside of cyberspace: cheating made easy. *Journal of Social Work Education* 35(3), 367-376. Retrieved from <http://www.jstor.org/stable/23043562>
- Gipp, B., Meuschke, N., & Breitingner, C. (2014). Citation-based plagiarism detection: Practicability on a large-scale scientific corpus. *Journal of the Association for Information Science and Technology*. 65(8), 1527-1540. doi: 10.1002/asi.23228
- Gu, Q., & Brooks, J. (2008). Beyond the accusation of plagiarism. *System* 36(1), 337-352. doi: 10.1016/j.system.2008.01.004
- Graham-Matheson, L., & Starr, S. (2013). Is it cheating-or learning the craft of writing? Using Turnitin to help students avoid plagiarism. *The Journal of the Association for Learning Technology*, 21(1), 1-13. doi: 10.3402/rlt.v21i0.17218
- Gullifer, J., & Tyson, G. A. (2010). Exploring university students' perceptions of plagiarism: a focus group study. *Studies in Higher Education*, 35(4), 463-481. doi: 10.1080/03075070903096508
- Hayes, N. & Introna, L. D. (2005). Cultural values, plagiarism, and fairness: When plagiarism gets in the way of learning. *Ethics and Behavior*, 15(3), 213-231. doi: 10.1207/s15327019eb1503_2
- Halliwell, S. (2002). The aesthetics of mimesis: Ancient texts and modern problems. Princeton University Press, New York.
- Hannabuss, S. (2001). Contested texts: issues of plagiarism. *Library Management*, 22(7), 311-318. doi: 10.1108/EUM0000000005595
- Heckler, N., Rice, M., & Hobson, B. (2013). Turnitin systems: A deterrent to plagiarism in college classrooms. *JRTE*, 45(3), 229-248. Retrieved from <http://go.galegroup.com/>

ps/i.do?id=GALE%7CA349903224&v=2.1&u=ko_acd_uoo&it=r&p=AONE&sw=w&asid=21e8d8faa8cd78143d557a0617932be8

- Hill, J.D., & Fetyko-Page, E. (2009). An empirical research study of the efficacy of two plagiarism-detection applications. *Journal of Web Librarianship*, 3(1), 169-181. doi: 10.1080/19322900903051011
- Howard, R. (1995). Authorships, and the academic death penalty. *College English*, 57(7), 788-806. doi: 10.2307/378403
- Howard, R. (2000). Sexuality, Textuality: The cultural work of plagiarism. *College English*, 62(4), 473-491. doi: 10.2307/378866
- Howard, R. (2007). Understanding internet plagiarism. *Computers and Composition*, 24(1), 3-15. doi: 10.1016/j.compcom.2006.12.005
- Howard, R. (2010). The scholarship of plagiarism: Where we've been, where we are, what's needed next. *Writing Program Administration*. 33(3), 117-124. Retrieved from http://go.galegroup.com/ps/i.do?id=GALE%7CA242454043&v=2.1&u=ko_acd_uoo&it=r&p=AONE&sw=w&asid=23c8e45bd5852bf97bb6c2d879298f12
- Hunt, J., & Tompkins, P. (2014). A comparative analysis of SafeAssign and Turnitin. *The Journal of the Virginia Community Colleges*. 19(6), 62-73. Retrieved from <http://commons.vccs.edu/cgi/viewcontent.cgi?article=1012&context=inquiry>
- Jenson, J., & de Castell, S. (2004). 'Turn It In': technological challenges to academic ethics. *Education, Communication & Information*, 4(2/3). 88-97. doi: 10.1080/14636310412331304735
- IEEEExplore. (2016, March 23) Digital Library. Retrieved from <http://about.jstor.org/10things>

- Iyer, P., & Singh, A. (2005). Document similarity analysis for a plagiarism detection system. Paper presented at 2nd Indian International Conference on Artificial Intelligence, Tumker, India.
- JSTOR. (2016, March 21) New to JSTOR. Retrieved from <http://about.jstor.org/10things>
- Kakkonen, T., & Mozgovoy, M. (2010). Hermetic and web plagiarism detection systems for student essays – an evaluation of the state-of-the-art. *Educational Computing Research*, 42(2), 135-159. doi: 10.2190/EC.42.2.a
- Klein, D. (2011). Why learners choose plagiarism: A review of literature. *Interdisciplinary Journal of E-Learning and Learning Objects*, 7(1), 97-110.
- Kutz, E., Rhodes, W., Sutherland, S., & Zamel, V. (2011). Addressing plagiarism in a digital age. *Journal of Sociology of Self-Knowledge* 9(3), p.15. Retrieved from <http://search.proquest.com.uproxy.library.dc-uoit.ca/docview/920098400?accountid=14694>
- Liddell, Jean (2003). A comprehensive definition of plagiarism. *Community & Junior College Libraries*, 11(3), 42-53. doi: 10.1300/J107v11n03_07
- Liddell, J., & Fong, V. (2005). Faculty perceptions of plagiarism. *Journal of College and Character*, 6(2). doi: 10.1080/02763915.2011.591709
- Lofstrom, E., & Kupila, P. (2013). The instructional challenges of student plagiarism. *Journal of Academic Ethics*, 11(3), 231-242. doi: 10.1007/s10805-013-9181-z
- Malloch, A. (1976). A dialogue on plagiarism. *College English*, 38(2), 165-174. Retrieved from <http://www.jstor.org/stable/376341>
- Maurer, H., Kappe, F., & Zaka, B. (2006). Plagiarism – A survey. *Journal of Universal Computer Science*, 12(8), 1150-1184. doi: 10.3217/jucs-012-08-1050

- Marsh, B. (2004). Turnitin.com and the scriptural enterprise of plagiarism detection. *Computers and Composition* 21(1), 427-438. doi: 10.1016/j.compcom.2004.08.002
- McGill, M. (2013). Copyright and intellectual property: The state of the discipline. *Book History*, 16(1), 387-427. doi: 10.1353/bh.2013.0010
- Oghigian, K., Rayner, M. & Chujo, K. (2016). A quantitative evaluation of Turnitin from a l2 science and engineering perspective. *CALL-EJ* 17(1), 1-18. Retrieved from http://callej.org/journal/17-1/Oghigian_Rayner_Chujo2016.pdf
- Park, C. (2004). Rebels without a clause: towards an institutional framework for dealing with plagiarism by students. *Journal of Further and Higher Education*, 28(3), 291-306. doi: 10.1080/0309877042000241760
- Park, C. (2003). In other (people's) words: Plagiarism by university students - literature and lessons. *Assessment & Evaluation in Higher Education*, 28(5), 471-488. doi: 10.1080/02602930301677
- Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *TESOL Quarterly*, 30(2), 201-230. doi: 10.2307/3588141
- Price, M. (2002). Beyond "gotcha": Situating plagiarism in policy and pedagogy. *College Composition and Communication*, 54(1), 88-115. doi: 10.2307/1512103
- Questia. (2016, March 25) Explore our library. Retrieved from <https://www.questia.com/#/library>
- Reymen, J. (2010). The rhetoric of intellectual property: Copyright law and the regulation of digital property. New York: Routledge.

- Rogers, S. (2009). The effects of plagiarism detection services on the teacher-student relationship as it pertains to trust. (Unpublished doctoral dissertation). Northern Illinois University, DeKalb, Illinois.
- Sabapathy, E., & Rahim, R. (2009). Realigning the focus of plagiarism detection using plagiarismdetect.com. *English Language Teaching*, 2(4), 210-214.
- Sage Publishing. (2016, Feb 17) Company Information. Retrieved from <https://us.sagepub.com/en-us/nam/company-information>
- Salem, R., & Bowers, W. (1970). Severity of formal sanctions as a deterrent to deviant behavior. *Law and Society Review*, 5(1), 21-40. Retrieved from <http://www.jstor.org/stable/3053071>
- ScienceDirect. (2016, March 5) *Gale serves the library community*. Retrieved from <http://solutions.cengage.com/gale/about/>
- Shi, L. (2012a). Common knowledge, learning, and citation practices in university writing. *Research in the Teaching of English*, 45(3), 308-334. Retrieved from <http://www.jstor.org.uproxy.library.dc-uoit.ca/stable/40997768>
- Shi, L. (2012b). Rewriting and paraphrasing source texts in second language writing. *Journal of Second Language Writing* 21(3), 134-148. doi:10.1016/j.jslw.2012.03.003
- Smiers, J. (2010). Rosemary Coombe, *The cultural life of intellectual properties: Authorship, appropriation, and the law*. *International Journal of Cultural Policy*, 16(1), 78-79. doi: 10.1080/10286630902989043
- Springer Publishing. (2016, March 25) About Us. Retrieved from <http://www.springerpub.com/about-us/>
- Stapleton, P. (2012). Gauging the effectiveness of anti-plagiarism software: An empirical study

- of second language graduate writers. *Journal of English for Academic Purposes*, 11(1), 125-133. doi: 10.1016/j.jeap.2011.10.003
- Sutherland-Smith, & W., Carr, R. (2005). Turnitin.com: Teachers' perspectives of anti-plagiarism software in raising issues of educational integrity. *Journal of University Teaching & Learning Practice*, 2(3), 92-101. Retrieved from <http://ro.uow.edu.au/jutlp/vol2/iss3/10/>
- Swagerman, S. (2008). The scarlet p: plagiarism, panopticism, and the rhetoric of academic integrity. *College Composition and Communication*, 59(4), 676-710. Retrieved from <http://www.jstor.org.uproxy.library.dc-uoit.ca/stable/20457030>
- Taylor & Francis Group (2016, March 21). *About Taylor & Francis Group*. Retrieved from <http://taylorandfrancis.com/about/>
- Turnitin. (2015, January 21) Our company. Retrieved from http://turnitin.com/en_us/about-us/our-company
- Turnitin. (2016, March 5) Originality check. Retrieved from https://guides.turnitin.com/01_Manuals_and_Guides/Instructor/Instructor_User_Manual/21_Originality_Check
- Turnitin. (2016, March 5) Content. Retrieved from http://turnitin.com/en_us/what-we-offer/originality-checking?layout=edit&id=121
- Thompsett, A., & Ahluwalia, J. (2010). Students turned off by Turnitin? Perception of plagiarism and collusion by undergraduate bioscience students. *Bioscience Education*, 16(3), 1-16. doi: 10.3108/beej.16.3
- Uzuner, O., Katz, B., & Nahnsen, T. (2005). Using syntactic information to identify plagiarism. Paper presented at Proceedings of the 2nd workshop on building educational applications using NLP, Ann Arbor, Michigan.

- Vie, S. (2013). A pedagogy of resistance toward plagiarism detection technologies. *Computers and Composition*, 30, 3-15. doi: 10.1016/j.compcom.2013.01.002
- Wheeler, D., & Anderson, D. (2010). Dealing with plagiarism in a complex information society. *Education, Business and Society: Contemporary Middle Eastern Issues*, 3(3), 166-177. doi: 10.1108/17537981011070082
- Williams, B.T., (2008). Trust, betrayal and authorship: Plagiarism and how we perceive students. *Journal of Adolescent & Adult Literacy*, 51(4), 350-354. doi: 10.1598/JAAL.51.4.6
- Wright, M. (1989). Plagiarism and the news. *Journal of Mass Media Ethics: Exploring Questions of Media Morality*, 4(2), 265-280. doi: 10.1080/08900528909358348

Appendix A – Results of Exploratory Studies

Article by Published Year	Author	Similarity Index (percentage score)	Primary Sources	Comments
1. 1970 – “Severity of Formal Sanctions as a Deterrent to Deviant Behavior”	Salem & Bowers, 1970	<ul style="list-style-type: none"> • 16% textual content matched • 84% textual content not matched 	<ul style="list-style-type: none"> • Majority of original/primary written submitted content not matched by Turnitin. • Some textual content including references and partial matching sentences found in academic articles sourced through a number of publishers including Wiley Online, Taylor and Francis Online and Sage Journals • Remainder of matching/similar content found in several student papers. This included shared references and parts of sentences. 	<ul style="list-style-type: none"> • Turnitin was successful at matching some text (16%) to a variety of sources, including student papers and academic articles found online, however, the software did not identify the original primary document source. • For the purpose of the Exploratory study the sample was originally sourced from JSTOR.org, a subscription-based digital library sourcing academic literature. Turnitin was not able to identify the original source material from JSTOR. • Possible factors contributing to this include: 1) Due to date of publication, a greater likelihood that the submitted article is not widely available in existing online journal databases; 2) The article may only be available in password protected online databases that Turnitin does not currently have access to; 3) Given the age of the article, it may not have been cited as regularly in academic work, such as more recently published journal articles or student papers catalogued by Turnitin, and therefore not as easily detected by the software. • It is worth noting that all three articles, where the majority of original content was not matched to a primary source (including this submission), were originally sourced through JSTOR.com by way of an article search through the University of Ontario Institute of Technology digital library service.
2. 1976 – “A Dialogue on Plagiarism”	Malloch, 1976	<ul style="list-style-type: none"> • 6% textual content matched • 94% textual content not matched 	<ul style="list-style-type: none"> • Majority of original/primary written submitted content not matched by Turnitin • 3% submitted material matched textual content found on scholarlywriting.net, including several sentence parts and a shared quotation (website is now inactive) • Some shared references/quotations (3% of 	<ul style="list-style-type: none"> • Turnitin was successful at matching some text (6%) to a variety of sources including a number of Internet sources/websites and academic articles found online, however, the software did not identify the original primary document source. • For the purpose of the Exploratory study the sample was originally sourced from JSTOR.org, a subscription-based digital library sourcing academic literature. • Possibilities for this include: 1) Due to the date of publication, a greater likelihood that the submitted article is not widely available in existing online journal databases that Turnitin does not currently have access to; 2) The article may only be available in password protected online databases; 3) Given the age of article, it may not have been cited as regularly in academic work, such as more recently published journal articles or student papers catalogued by Turnitin, and therefore not as easily detected by the software. • It is worth noting that the three articles where majority of original content was not

			submitted material) sourced to student essays. This matching content included sentence parts. The details of which are not made available due to Turnitin's privacy policy around student submissions.	matched to a primary source were originally sourced to JSTOR.com by way of an online article search through the University of Ontario Institute of Technology digital library service.
3. 1982 – “The Language Game – Plagiarism: The Unfun Game”	Caroll, 1982	<ul style="list-style-type: none"> • 7% textual content matched • 93% textual content not matched 	<ul style="list-style-type: none"> • Majority of original/primary written submitted article content not matched by Turnitin. • 3% submitted matching textual material was found in several student papers catalogued by Turnitin (specific details were not available due to Turnitin's privacy restrictions) • Several (2%) shared references found in a single academic article sourced through Taylor and Francis Online (a publisher of Academic literature) • Remainder of matching textual content (2%) including shared references, quotations, and sentence parts, came from websites. 	<ul style="list-style-type: none"> • Turnitin was successful at matching some text (7%) to a variety of sources, including student papers, and academic articles found online, however, the software did not identify the original primary document source. • For the purpose of the Exploratory study the sample was originally sourced from JSTOR.org, a subscription-based digital library sourcing academic literature. • Possibilities for this include: 1) Due to the date of publication, a greater likelihood that the submitted article is not widely available in existing online journal databases; 2) The article may only be available in password protected online databases; 3) Given the age of article, it may not have been cited as regularly in academic work, and therefore not as easily detected in other works such as published journal articles or student papers catalogued by Turnitin, and therefore not as easily detected by the software. • It is worth noting that the three articles where the majority of original content was not matched to a primary source were originally sources to JSTOR.com by way of an online article search through the University of Ontario Institute of Technology digital library service.
4. 1989 – “Plagiarism and the News Media”	Wright, 1989	<ul style="list-style-type: none"> • 99% • 1% textual content unmatched 	<ul style="list-style-type: none"> • Majority of original/primary submitted textual material (full article) found in Taylor and Francis Online (a publisher of academic journals and books). • Some content (under 2%) matched to several student 	<ul style="list-style-type: none"> • Turnitin was successful in matching submitted textual content to an existing primary source (the original academic journal article published online). • Matching content found in article published in online database. • Unmatched textual material appeared to be random sampling of words throughout document (1%). • Turnitin also successfully matched some content to several Internet sites and student papers. • This content consisted primarily of references and sourced definitions including direct

			papers Matching content included references and sentence parts. <ul style="list-style-type: none"> Some matching content found on several Internet sites (4%) including original content/parts of sentences and shared references. 	quotations although copies of student papers are not provided in full due to privacy restrictions. <ul style="list-style-type: none"> References to submitted article appeared to be accurate (although specific details surrounding sourcing in student papers were not available due to privacy restrictions).
5. 1994 – “Scientific Misconduct and Editorial and Peer Review Processes”	Fox, 1994	<ul style="list-style-type: none"> 95% textual content matched 5% textual content unmatched 	<ul style="list-style-type: none"> Majority of original/primary textual material (full article) found in Gale Publishing. Partial preview of article found in Questia (an online research and paper writing service) 	<ul style="list-style-type: none"> Turnitin was successful in matching the submitted textual content to an existing primary source (the original academic journal article published online). Article was sourced to two publishing databases including one full copy available in Gale Publishing and a Preview of article found in Questia Publishing. Unmatched textual material included text describing the search terminology from the original pdf version of the original article download (5%). Some matching material was found in shared references, including citations of original paper. These references appeared to be accurate. Secondary matching material included shared quotations around definitions of plagiarism.
6. 1999 – “The downside of cyberspace: cheating made easy”	Gibelman et al., (1999)	<ul style="list-style-type: none"> 100% textual content matched 	<ul style="list-style-type: none"> Majority of original/primary textual material (full article) found in Gale Publishing and the Ontario Ministry of Education (OSAPAC) A preview of the article was matched to Questia (an online research and paper writing service) Small selections of content found (including partial sentences and shared references) on a variety of online sites and student essays sourced by Turnitin. 	<ul style="list-style-type: none"> Turnitin was successful in matching the submitted textual content to an existing primary source (the original academic journal article published online). Content found in article published in online database (Gale Publishing).
7. 1999 – “Internet Plagiarism:	Austin & Brown, 1999	<ul style="list-style-type: none"> 99% textual content 	<ul style="list-style-type: none"> Majority of original/primary textual material (full article) found in 	<ul style="list-style-type: none"> Turnitin was successful in matching the submitted textual content to an existing primary source (the original academic journal article published online). Article sourced to one online publisher and two university databases.

Developing Strategies to Curb Student Academic Dishonesty”		<p>matched</p> <ul style="list-style-type: none"> • 1% textual content unmatched 	<p>two academic databases (University of Florida and McMaster University) and Science Direct</p> <ul style="list-style-type: none"> • Small selections (under 3%) matching material found in student papers (ex. references, partial sentences etc.) 	<ul style="list-style-type: none"> • Unmatched textual material appears to be random sampling of text throughout document. • Primary content found in three places overall. • Turnitin also matched textual content to several student papers. This content included a single shared reference with the original submission.
8. 2009 – “Realigning the Focus of Plagiarism Detection Using Plagiarismdetect.com”	Sabapathy et al., 2009	<ul style="list-style-type: none"> • 100% textual content matched 	<ul style="list-style-type: none"> • Majority of original/primary textual material (full article) found in Canada Centre of Science and Education database (publisher of academic journals) • Selections of textual content (8% and less) found in several student papers (matches consisted of similarities in original written content and references) • Also small content matches (4% or less) with several academic articles, including content and references) 	<ul style="list-style-type: none"> • Turnitin was successful in matching the submitted textual content to an existing primary source (the original academic journal article published online). • Significant amount of textual content also matched to several student papers, where the article was commonly referenced, as well as in several published academic articles. (commonly referenced article). • More recent publication date may provide easier access to original article as more likely to be available in searchable online publishing databases.
9. 2010 - “Plagiarism Detection Services: How well do they actually perform?”	Fiedler & Kaner, 2010	<ul style="list-style-type: none"> • 93% textual content matched • 7% textual content unmatched 	<ul style="list-style-type: none"> • Majority of original/primary textual material (full article) found on author’s (C. Kaner) personal website. • A single reference from the article was also sourced to IEEEExplore Digital Library, a technology-based academic journal offering as well as Turnitin.com 	<ul style="list-style-type: none"> • Turnitin was successful in matching the submitted textual content to an existing primary source – a full copy of the article available on the author’s personal website. • Unmatched textual material appears to be random sampling of text and table/chart results (7%). • More recent publication date may provide easier access to original article as more likely to be available in searchable online publishing databases.
10. 2013 – “An analysis of student privacy	Brinkman, 2013	<ul style="list-style-type: none"> • 99% textual content 	<ul style="list-style-type: none"> • Majority of original/primary textual material (full article) found in 	<ul style="list-style-type: none"> • Turnitin was successful in matching the submitted textual content to an existing primary source (the original academic journal article published online).

rights in the use of plagiarism detection software”		<p>matched</p> <ul style="list-style-type: none"> • 1% textual content unmatched 	<p>Springer Publishing</p> <ul style="list-style-type: none"> • 10% of article content found in a variety of student papers including shared references sourced through Turnitin database (specific details of these matches were not available due to Turnitin privacy policy) • Small selections of textual content (1% or less) found in a variety of Internet sources (see websites, online publications) 	<ul style="list-style-type: none"> • Also demonstrated ability to match similar textual content in student papers and on a variety of Internet sources. • More recent publication date may provide easier access to original article as more likely to be available in searchable online publishing databases. • Unmatched textual material appears to be random occurrence of unmatched text (1%).
---	--	---	---	--